

⑨ 日本国特許庁(JP)

⑩ 特許出願公開

⑫ 公開特許公報(A)

昭60-200300

⑤ Int.Cl.⁴

G 10 L 5/06

識別記号

庁内整理番号

8221-5D

④ 公開 昭和60年(1985)10月9日

審査請求 未請求 発明の数 1 (全9頁)

⑭ 発明の名称 音声の始端・終端検出装置

⑮ 特 願 昭59-56622

⑯ 出 願 昭59(1984)3月23日

⑰ 発 明 者 森 井 秀 二 川崎市多摩川区東三田3丁目10番1号 松下技研株式会社内

⑱ 発 明 者 藤 井 諭 川崎市多摩川区東三田3丁目10番1号 松下技研株式会社内

⑲ 発 明 者 星 見 昌 克 川崎市多摩川区東三田3丁目10番1号 松下技研株式会社内

⑳ 出 願 人 松下電器産業株式会社 門真市大字門真1006番地

㉑ 代 理 人 弁理士 中尾 敏男 外1名

明 細 書

1、発明の名称

音声の始端・終端検出装置

2、特許請求の範囲

(1) 音声を含む信号から一定時間長の区間毎に信号のエネルギーとスペクトル形状を表わす特徴量を抽出する特徴量抽出部と、前記特徴量を用いて入力された信号が有音であるか無音であるか一定時間長の区間毎に判定する有音・無音判定部と、前記有音・無音の判定結果の時系列を用い判定結果の持続時間により音声の始端・終端の候補を検出する始端・終端候補検出部と、始端・終端候補の前後における信号のエネルギー変化とスペクトルの変化の大きさを用いて始端・終端の位置を決定する始端・終端決定部とを具備することを特徴とする音声の始端・終端検出装置。

(2) 信号のスペクトル形状を表わす特徴量として、帯域フィルタ群、フーリエ変換若しくは線形予測分析のいずれかの方法により求められるパワ

ースペクトルまたは線形予測分析により得られるLPCKエプストラム係数のいずれかを用いることを特徴とする特許請求の範囲第1項記載の音声の始端・終端検出装置。

(3) 有音・無音判定部が、信号のエネルギーと二つの閾値との比較を行う第1判定部と、無音、無声音、有声音の三つの標準パターンと入力信号のスペクトルとの統計的距離尺度を用いスペクトルの類似度による判定を行う第2判定部とを備え、前記統計的距離尺度として線形判別関数、マハラノビス距離、ベイズ判定のいずれかを用いることを特徴とする特許請求の範囲第1項記載の音声の始端・終端検出装置。

(4) 始端・終端決定部のスペクトルの変化の大きさを表わす特徴量として、一定時間長の区間におけるスペクトルを表わす特徴量と前の区間のスペクトルを表わす特徴量とのユークリッド距離を用いることを特徴とする特許請求の範囲第1項記載の音声の始端・終端検出装置。

3、発明の詳細な説明

産業上の利用分野

本発明は音声認識装置に用いられる音声の始端・終端の検出装置に関するものである。

従来例の構成とその問題点

音声の始端、終端の検出方法に関する従来例としては、信号のエネルギーと零交差回数を用いた方法が知られている。これは、新美康永：音声認識、共立出版（1979）、あるいは、

L. R. Rabiner and M. R. Sambur: An algorithm for determining the endpoint of isolated utterances, Bell Syst. Tech. J., (1975) に示されている。

零交差回数というのは信号の符号のみを残し、振幅を1ビットに量子化した零交差波の一定時間長の区間における零交差の平均回数である。音声のようにスペクトル構造をもった信号の零交差回数はスペクトル中の優勢な周波数成分とよく対応する。第1図(a)~(c)は音声信号の零交差回数の分布を示したもので、(a)は無音、(b)は無声音、(c)は有声音の分布である。図から分るように、音声信

号の零交差回数は、有声音のように低域の周波数成分の優勢は音声では第1図(c)のように小さな値を示し、無声音のように高域の周波数成分の優勢な音声では第1図(b)のように大きな値を示す。従来法による音声の始端・終端検出方法はこの零交差回数を利用することにより信号のエネルギーは小さいが、零交差回数は大きな値をとる無声音の検出精度を上げた方法である。

以下図面を参照しながら従来例の音声の始端・終端検出方法について説明する。

第2図は従来例の構成を示したものであり、第3図は従来例における音声の始端・終端検出方法の動作を説明するための例を示したものである。音声を含む信号は第2図に示すエネルギー算出部1と零交差回数算出部2によりフレーム（例えば10 msec 長）毎に信号エネルギー $E(n)$ （ n はフレーム番号）と零交差回数 $Nz(n)$ という2つの特徴パラメータに変換される。3は信号のエネルギーレベルにより確実に音声区間であるという部分を検出する始端・終端候補決定部であり、信号エ

ネルギー $E(n)$ に対し2つの閾値 E_1 , E_2 ($E_1 > E_2$) を適用し音声の始端候補 n_1 、終端候補 n_2 を求める。これは第3図(a)の例に示すように、エネルギーの値が E_2 を越え、かつその後 E_2 以下になることなしに E_1 を越えるとき、音声区間に入ったとみなし、 E_2 を越えた点を始端候補 n_1 とするものである。終端候補 n_2 は時間軸を逆にして、同様の方法で決定する。第2図の4は音声の始端・終端決定部である。ここでは、零交差回数算出部2で計算された信号の零交差回数 $Nz(n)$ と閾値 N_0 を用いて、エネルギー $E(n)$ は小さいが零交差回数 $Nz(n)$ が大きな値をとる無声音が、始端・終端候補決定部3で定められた音声の始端・終端候補 (n_1 , n_2) の外側にないか検査する。第3図(b)の例に示すように、始端候補 n_1 より前の数フレームの区間において零交差回数 $Nz(n)$ が閾値 N_0 より大となるフレームの数を数え、その数が一定値（たとえば3）以上であれば始端候補 n_1 より前に無声音があるとみなし最初に閾値 N_0 を越えたフレーム n_1 に始端を移す。終端

についても同様である。ただし、第3図(b)では終端 n_2 はもとのままである場合を示している。このようにして最終的な音声の始端・終端 (n_1 , n_2) が決定される。

しかし、上記のように零交差回数を用いた方法では、エネルギーが小さく零交差回数も小さい有声音（例えば、/b/, /d/）などの脱落を減少することはできない。また、音声の始端、終端には唇を開けたときの雑音とか呼吸音による雑音が付加しやすい。第4図(a), (b)は上記雑音が付加した音声のエネルギー変化を示したもので、(a)は唇の動きによる雑音が始端に付加した場合の例として異様（/ijoo/）という音声のパワー変化を示し、(b)は呼吸音による雑音が始端に付加した場合の例として出湯（/ideju/）という音声のパワー変化を示したものである。図に示した例のような場合、従来例では始端は雑音部分になってしまう。このように、従来例による方法では始端、終端の位置を誤ってしまい音素の脱落や雑音による音素の付加が避けられない場合があるという欠

点がある。

発明の目的

本発明は上記欠点に鑑み、音声の脱落、雑音の付加が少なく、位置精度の高い音声の始端、終端検出装置を提供するものである。

発明の構成

上記目的を達成するためには、信号のエネルギーとスペクトル形状によりフレーム毎（例えば10 msec）に有音・無音の判定を行なう有音・無音判定部と、フレーム毎の有音・無音判定結果の持続性により音声の始端・終端候補を検出する部分と、無音から有音またはその逆の有音から無音に変化する場合における信号のエネルギーの変化とスペクトルの変化の大きさという動的な特徴により始端・終端の位置を決定する部分とを備え、入力された音声を含む信号から音声の始端・終端の位置を検出するようにしたものである。

実施例の説明

以下、本発明の実施例について図面を参照しながら説明する。

一の2つの閾値 E_1 、 E_2 ($E_1 > E_2$) が格納してある。また、標準パターンメモリ10には、無音・無声音を判別するための線形判別関数と無音・有声音を判別するための線形判別関数の2種類の線形判別関数の係数が格納されている。そして、これら2つの閾値 E_1 、 E_2 と2つの線形判別関数の係数は、あらかじめ使用する環境下で発声された音声データの統計処理により求められ、格納されている。11は始端・終端候補検出部であり、有音・無音判定部8より送られてくるフレーム毎の有音・無音判定結果の持続時間により、音声の始端・終端候補を検出する。12は始端・終端決定部で、最終的な始端・終端を決定する。なお、第6図8～12はマイクロプロセッサ1台で構成される。

以上のように構成された音声の始端・終端検出装置についてその動作を説明する。

マイク等より入力される音声を含む信号は第6図のエネルギー抽出部6およびスペクトル形状抽出部8によりフレーム毎にパワー PW と3つの帯域

第6図は本発明の一実施例における音声認識装置に組込まれた音声の始端・終端検出装置のブロック図を示したものである。図において6はエネルギー抽出部で、整流平滑回路で構成され信号のパワーをフレーム毎に抽出する。8はスペクトル形状抽出部で、例えば、低域(250～600Hz)、中域(600～1500Hz)、高域(1500～4000Hz)の3種類の帯域通過フィルタ群と整流平滑回路で構成され、各帯域におけるフレーム毎のパワーがスペクトル情報として用いられている。エネルギー抽出部6とスペクトル形状抽出部8とで特徴量抽出部13を構成する。7はマルチプレクサで、エネルギー抽出部6からの信号のパワーとスペクトル形状抽出部8からの帯域フィルタパワーを時分割で有音・無音判定部8へ入力するためのものである。8は有音・無音判定部で、無音、無声音、有声音の判別を行うためのものである。9、10は閾値メモリと標準パターンメモリであり有音・無音判定部8で用いられる定数値が格納されている。閾値メモリ9には、パワ

ワー P_i ($i=1\sim3$) に変換される。この PW 、 P_i はマルチプレクサ7を経て有音・無音判定部8に入力される。有音・無音判定部8では入力された PW 、 P_i ($i=1\sim3$) の4つのパラメータを対数変換し対数パワー LPW と対数帯域パワー LP_i ($i=1\sim3$) を求める。そして、 LPW と LP_i ($i=1\sim3$) の4つのパラメータと閾値メモリ9と標準パターンメモリ10に格納されている閾値 E_1 、 E_2 と2つの線形判別関数の係数とを用いて、入力されたフレームが有音であるか無音であるかを判定する。この有音・無音判定はまず最初に2つのエネルギー閾値 E_1 、 E_2 ($E_1 > E_2$) と対数パワー LPW との比較による判定が行なわれる。2つの閾値 E_1 、 E_2 は $LPW > E_1$ ならば確実に有音であり、 $LPW < E_2$ ならば確実に無音であるという値に設定されているため判定結果は式(1)に示すようなものとなる。

$$\left. \begin{array}{ll} LPW > E_1 & \text{ならば 有音} \\ LPW < E_2 & \text{ならば 無音} \\ E_2 \leq LPW \leq E_1 & \text{ならば 不定} \end{array} \right\} \quad \text{式(1)}$$

LPWというエネルギー量を用いた判定で不定という判定結果を得た場合は、さらにスペクトル形状による有音・無音判定を行なう。これは、低域、中域、高域の3つの帯域の対数パワー LP_i

($i = 1 \sim 3$) をスペクトル形状を表わすパラメータとし、標準パターンメモリ10に格納してある2種類の線形判別関数の係数を用い判別関数の値を計算することにより有音・無音を判定するものである。この2つの線形判別関数のうち1つは無音/無声音を判別するためのものであり、もう1つは無音/有声音を判別するためのものである。線形判別関数 FX は式(2)に示すものであり、標準パターンメモリ10には式(2)の A_i ($i = 1 \sim 3$) と $\overline{LP_i}$ ($i = 1 \sim 3$) が無音/無声音、無音/有声音という2種類の線形判別関数毎に格納されている。

$$FX = \sum_{i=1}^3 A_i (LP_i - \overline{LP_i}) \quad \dots\dots \text{式(2)}$$

(ただし、 A_i は係数、 $\overline{LP_i}$ は平均値)

式(2)における A_i は2つのクラスの最適な判別を行なうように設定され2つのクラスの級内分散、

級間分散の比であるFisher比の最大化条件から求められる。本実施例において、式(2)の A_i および $\overline{LP_i}$ はあらかじめ使用環境下で発声された音声データの無音・無声音・有声音を統計処理して求められる。そして FX の値は入力が無音のとき負で、入力が無声音あるいは有声音のときは正の値をとるように設定してある。したがって、スペクトル形状による有音・無音判定は無音/無声音と無音/有声音の2つの線形判別関数を計算しいずれか一方でも正の値をとるならば有音、2つとも負の値ならば無音と判定する。このようにして得られたフレーム毎の有音・無音の判定結果は第6図の始端・終端候補検出部11に送られる。始端・終端候補検出部11ではフレーム毎に得られる有音・無音の判定結果の持続時間により音声の始端候補および終端候補を検出する。11の始端・終端候補検出部はマイクロプロセッサの2つのレジスタをカウンタとして用い、さらに比較演算機能を用いて構成される。そして、始端候補検出においては1つのカウンタだけを用い、終端候補検

出ではカウンタを2つとも用いている。第6図は始端候補検出のための処理の流れを示したものである。第6図は有音と判定されたフレームが6フレーム以上連続したときその先頭のフレームを始端候補とすることを示している。第6図の処理Iは有音フレームのカウンタ(第6図のCOUNT)、始端候補フレーム番号格納領域(第6図FRAMES)そして処理フレームポジション(第6図I)の初期化のためのリセットである。第6図処理IIは処理フレームポジションの更新である。処理IIは処理フレームが有音であるか無音であるかの比較による分岐である。処理しているフレームが有音である場合は有音フレームのカウンタ(COUNT)に1を加える(第6図処理III)。さらに、始端候補フレーム番号格納領域(FRAMES)が0にリセットされたままである場合は現在処理を行なっているフレームの番号(I)を格納する(処理IV)。処理Vでは有音フレームのカウンタが6になったかの判定を行なう。そして、カウンタが6以下の場合は処理IIに戻り、カウンタが6以上に

なった場合は始端候補が検出されたということで始端候補検出処理を終了する。処理が終了するまでの間に処理IIにおいて無音であるというフレームがあった場合は、処理IIにおいて有音フレームカウンタおよび始端候補フレーム番号格納領域はリセットされ処理はIIに戻る。有音フレームカウンタは無音フレームがあると処理IIによりリセットされるため有音が連続したフレーム数のカウンタとなる。したがって、処理IIの判定は有音が6フレーム以上連続したかの判定となる。したがって、音声の始端の前に唇の動きによる雑音などで有音と判定されたフレームが2~3フレームあったとしてもその後1フレームでも無音と判定されるフレームがあればそれは除去される。このようにして始端候補が検出されると次に終端候補検出のための処理が行なわれる。第7図は終端候補検出のための処理の流れを示したものである。

第7図の処理Iは無音フレームのカウンタ(第7図のCOUNT1)、有音フレームのカウンタ(第7図のCOUNT2)そして終端候補フレー

ム番号格納領域(第7図FRAME)の初期化のためのリセットである。第7図処理ロは処理フレームポジション(第7図I)の更新である。処理ハは処理フレームが有音であるか無音であるかの比較による分岐である。処理しているフレームが無音である場合は無音フレームカウンタを更新し、有音フレームカウンタをリセットする(処理ニ、ホ)。さらに無音カウンタが2以上でかつ終端フレーム番号格納領域がリセットされている場合には無音フレームカウンタが1となったフレームの番号を終端候補フレームとして終端フレーム格納領域に格納する(処理ヘ、ト)。処理チでは無音フレームカウンタが30になったかの判定を行なり。そして、無音フレームカウンタが30未満の場合は処理ロに戻り、30以上となった場合は音声を終了したとみなし処理を終了する。処理ハにおいて有音であった場合に分岐する処理リ、ヌ、ルは終端候補フレームが格納されてから有音のフレームが何フレーム連続したかの処理で5フレーム以上連続した場合は、音声は終了していな

いとみなし処理イに戻り終端候補検出をやり直す。有音フレームが5フレーム未満の場合は雑音とみなし、その区間は無音区間であるということで処理ニにおいて無音フレームカウンタにその区間長が加えられる。

終端候補は無音フレームが2フレーム連続したとき音声の終了の可能性があるとし先頭の無音フレームを音声の終了候補とし、その終了候補フレームから29フレームの間に有音フレームが5フレーム以上連続することがない場合は先の終了候補フレームを終端候補とする。もし、終了候補から29フレーム後の間に有音フレームが5フレーム以上連続した場合は、音声はまだ終了していないとし、カウンタおよび終了候補フレームをすべてリセットし第7図に示す終端検出処理を次のフレームからやり直す。このような処理により終端に付加された4フレーム以下の雑音は取り除かれる。始端・終端決定部12では始端・終端候補検出部11により検出された始端・終端候補フレーム付近におけるパワーLPWとスペクトルLPiの変化

の大きさにより最終的な始端・終端を決定する。パワーの変化の大きさを表わすパラメータとしては式(3)に示すようにフレーム毎に得られる対数パワーLPWの差分値LPWDが用いられる。

$$LPWD_j = LPW_j - LPW_{j-1} \quad \dots\dots \text{式(3)}$$

(ただし、jはフレーム番号)

また、スペクトルの変化の大きさを表わすパラメータとしては式(4)に示す帯域対数パワーLPiのユークリッド距離SPDを用いる。

$$SPD_j = \sum_{i=1}^3 (LPi_j - LPi_{j-1})^2 \quad \dots\dots \text{式(4)}$$

(ただし、iは帯域を表わし、jはフレーム番号を表わす)

LPWDというパラメータはパワーが増加している場合正の値をとり、パワーが減少している場合は負の値をとる。また、SPDは無音から有音へと変化する場合のようにスペクトルの形状が大きく変化するところでは大きな値をとる。始端の決定はまず始めにLPWDが正の値をとるフレームを終端候補から後端に向かって検索する。次にLPWDが最初に正となったフレームから後2フ

レームの計3フレームの中でLPWDが正の値でSPDが最大となるフレームを求め、そのフレームを終端フレームと決定する。

終端の決定は、まず始めにLPWDが負の値をとるフレームを終端候補フレームから始端方向に向かって検索する。次にLPWDが最初に負となったフレームから2フレーム前の計3フレームの中でLPWDが負の値でSPDが最大となるフレームを求め、そのフレームの1つ前のフレームを終端フレームと決定する。このようにして得られた始端・終端は音声認識装置にて利用される。

本実施例によれば、有音・無音判定部8においてエネルギーレベルが低い入力信号に対し、線形判別関数を用い無音とのスペクトル形状の相異により有音であるか無音であるか判定する方法をとっているため、エネルギーの小さな無声子音や有声子音の脱落を減少することができる。また、始端・終端候補検出部11において、音声の持続性を考慮した検出を行なっているため、音声の始端・終端前後に付加された短かい雑音を取り除くこと

ができる。さらに、始端・終端決定部12では、無音から有音あるいは逆の場合におけるエネルギーの変化とスペクトル形状の変化の大きさを利用して始端・終端の位置を決定しているため位置精度の高い音声の始端・終端を得ることができる。第8図は「土台」(ノドダイノ)と発声された音声に本発明の一実施例における始端・終端検出を適応した例で、第8図aは対数パワーLPWを示し、bはスペクトル変化SPD、cはパワー変化LPWD、dの実線は無音/無声音を判別する線形判別関数の値、破線は無音/有声音を判別する線形判別関数の値を示したものである。第8図の例においては、始端・終端にそれぞれ雑音が見られる。フレーム毎の無音・有音判定部8では、LPWが E_1 以上であるか、またはLPWが E_1 と E_2 の間にある場合はdに示す二つの線形判別関数の正負を勘案することにより、aに示すイからロおよびハからニの区間を有音と判定する。これにより始端の雑音を取り除かれる。始端・終端候補検出部11においては、有音・無音フレームの

持続性により始端候補フレームをイとし、終端候補フレームをロとする。このときハからニの有音区間は6フレーム未満であるため雑音と判定される。そして、始端・終端決定部12では対数パワーの変化cとスペクトルの変化bにより始端イ、終端ロが決定され雑音が除去された正しい始端・終端の位置が得られる。あらかじめ目視による始端・終端のラベル付けが行なわれている男性話者1名が発声した212単語を用いて本発明の一実施例の評価実験を行なった結果、ラベルとの差が2フレーム以内となるものが始端で93.4%、終端で92.9%、ラベルとの差が3フレーム以内となるものが始端で97.6%、終端で97.2%という結果を得た。そして、始端の音素脱落という重大な誤りは2単語、終端の音素脱落という重大な誤りは2単語と少なく、また雑音の付加による誤りはなく、良好な結果を得ることができ、本発明による音声の始端・終端検出装置が有効に動作することを確かめることができた。

なお、以上の説明ではスペクトル形状を表わす

パラメータとして帯域対数パワーを用いた有音・無音の判定として線形判別関数を用いた場合について説明したが、スペクトル形状を表わすパラメータとして信号のフーリエ変換や線形予測分析により得られるパワースペクトルや線形予測分析により得られるLPC-ケプストラム係数を用い、有音・無音の判定法としてベイズ判定やマハラノビス距離などの統計的距離尺度を用いても良い。

発明の効果

以上のように、本発明は信号のエネルギー情報だけでなくスペクトル形状をも用いたフレーム毎の有音・無音判定部と、音声の持続性を考慮した始端・終端候補検出部と、エネルギーの変化およびスペクトル形状の変化量により始端・終端位置を決定する決定部により構成される音声の始端・終端検出装置を提供するもので、有音・無音判定として、無音・無声音・有声音のスペクトル標準パターンとの統計的距離尺度を用いたスペクトル形状の相異を利用しているため、エネルギーの小さな無声子音や有声子音の脱落を少なくでき、ま

た有音の持続性による始端・終端候補検出を行なっているため雑音の付加が少なく、しかもエネルギーとスペクトルの変化の大きさにより始端・終端の位置を決定するため位置精度が高いというすぐれた効果が得られる。

4、図面の簡単な説明

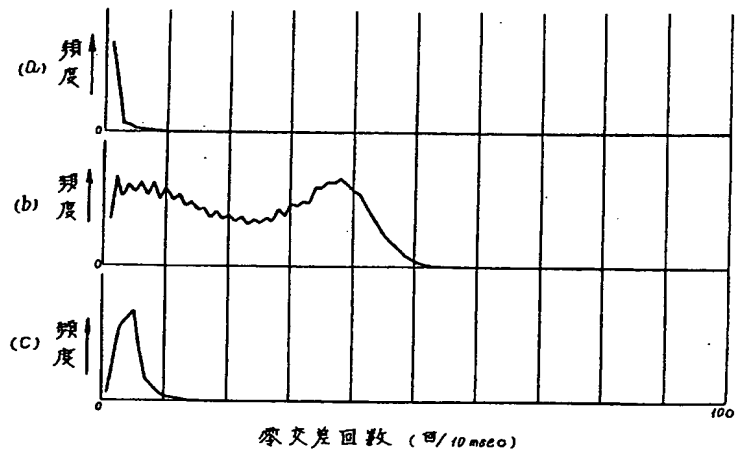
第1図は従来用いられている零交差回数分布図、第2図は従来の始端・終端検出装置のブロック図、第3図は従来の始端・終端検出装置の動作例を説明する図、第4図は従来で雑音が付加した音声のエネルギー変化を示す図、第5図は本発明の一実施例における音声の始端・終端検出装置のブロック図、第6図は本発明の一実施例における始端候補検出処理を示すフローチャート図、第7図は本発明の一実施例における終端候補検出処理を示すフローチャート図、第8図は本発明の一実施例における動作例を説明する図である。

5……エネルギー抽出部、6……スペクトル形状抽出部、7……マルチプレクサ、8……有音・無音判定部、9……閾値メモリ、10……標準バ

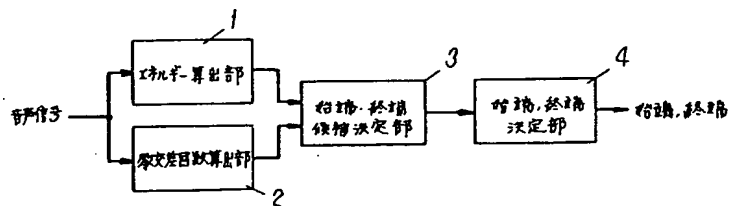
ターンメモリ、11……始端・終端候補検出部、
 12……始端・終端決定部、13……特徴量抽出
 部。

代理人の氏名 弁理士 中 尾 敏 男 ほか1名

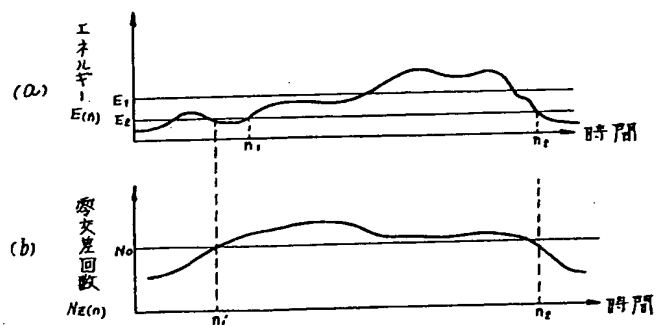
第 1 図



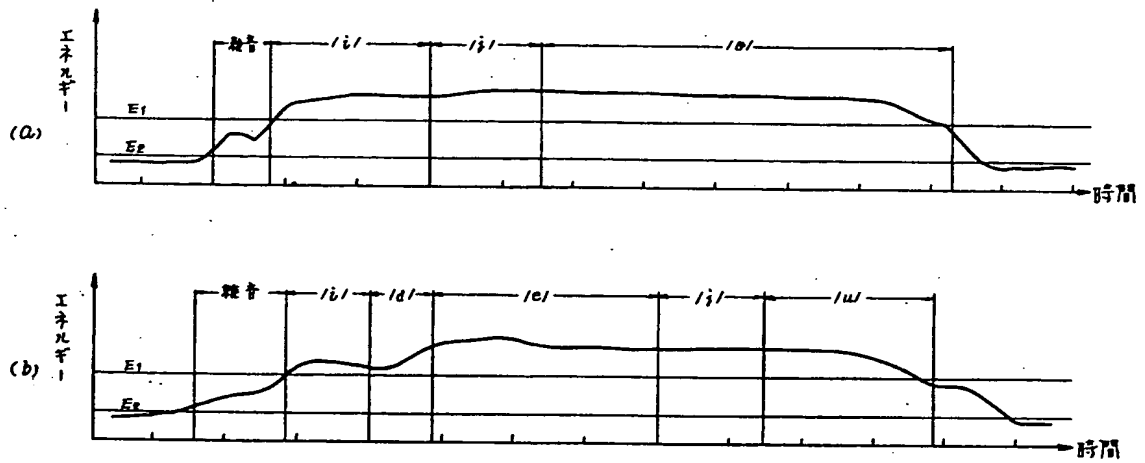
第 2 図



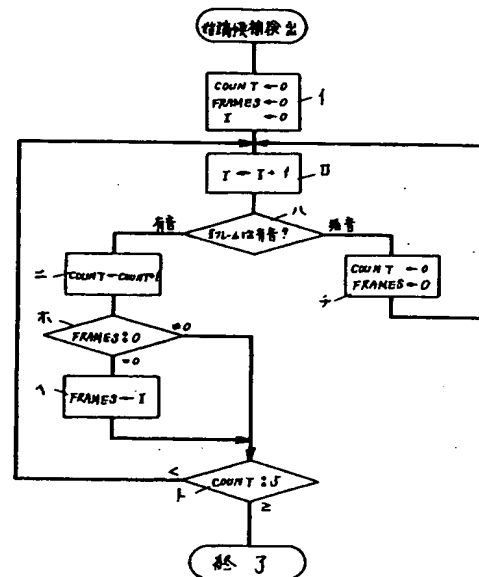
第 3 図



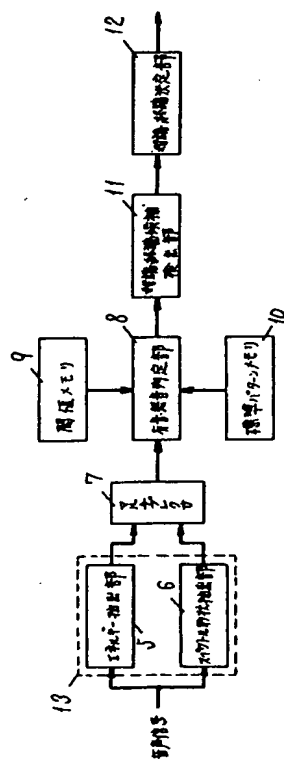
第 4 圖



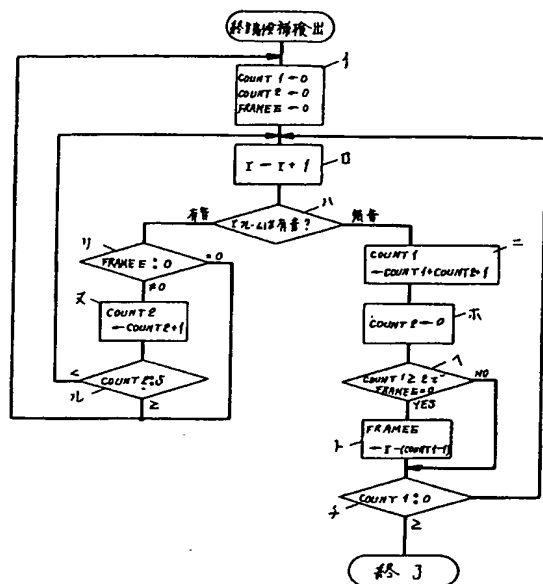
第 6 圖



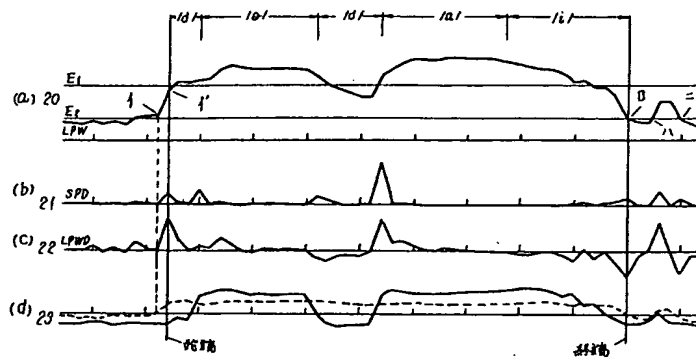
第 5 圖



第 7 図



第 8 図



| | |
|----------------|--|
| (72) Inventor | Hideji Morii in Matsushita Research Institute Co., Ltd. 3-10-1 Higashi Mita, Tanagawa-ku, Kawasaki-shi |
| (72) Inventor | Ron Fujii in Matsushita Research Institute Co., Ltd. 3-10-1 Higashi Mita, Tanagawa-ku, Kawasaki-shi |
| (72) Inventor | Masami Hoshimi in Matsushita Research Institute Co., Ltd. 3-10-1 Higashi Mita, Tanagawa-ku, Kawasaki-shi |
| (71) Applicant | Matsushita Electric Industrial Co., Ltd. 1006 Kadoma, Oaza, Kadoma-shi |
| (74) Agent | Toshio Nakao, patent attorney (and 1 other) |

[the notation [?] = poorly legible or illegible]

2. Claims

(1) A voice initiation-termination detection device that is characterized in that it has a feature quantity extraction unit that extracts from a signal that includes a voice a feature quantity that represents the energy of the signal and the shape of its spectrum in each section of a fixed duration; a sound-or-silence discrimination unit that uses said feature quantity to discriminate whether the signal that has been inputted is sounded or silent in each section of a fixed duration; an initiation-termination [?] unit that uses a time series of said sound-or-silence discrimination results to detect the [?] of the initiation-termination of the voice by the [?] of the discrimination results; and an initiation-termination decision unit that uses the size of the energy change and spectrum change of the signal before and after the initiation-termination [?] to decide the position of the initiation-termination.

(2) A voice initiation-termination detection device as described in claim 1 that is characterized in that it uses as the feature quantity that expresses the shape of the spectrum of the signal either the power spectrum determined by the method of band filters, Fourier transformation, or

linear prediction analysis, or the LPC [linear predictive coding] spectrum coefficient obtained by linear prediction analysis.

(3) A voice initiation-termination detection device as described in claim 1 that is characterized in that the sounded-silent discrimination unit has a first discrimination unit that makes a comparison between the energy of the signal and two regions[?] and a second discrimination unit that uses a statistical distance scale between three [?] patterns -- silent, voiced, and voiceless -- and the spectrum of the input signal and discriminates among them based on the degree of similarity of the spectrum, and in that as said statistical distance scale it uses either a linear discrimination function, the Mahalanobis distance, or Bayes discrimination.

(4) A voice initiation-termination detection device as described in claim 1 that is characterized in that it uses as the initiation-termination decision unit's feature quantity that represents the size of the spectrum change the Euclidean distance between the feature quantity that represents the spectrum in a section of fixed duration and the feature quantity that represents the spectrum of said section.